

# Multiple Wavelet Basis Image Denoising Using Besov Ball Projections

Hyeokho Choi, *Member, IEEE*, and Richard G. Baraniuk, *Fellow, IEEE*

**Abstract**—We propose a new image denoising algorithm that exploits an image’s representation in multiple wavelet domains. Besov balls are convex sets of images whose Besov norms are bounded from above by their radii. Projecting an image onto a Besov ball of proper radius corresponds to a type of wavelet shrinkage for image denoising. By defining Besov balls in multiple wavelet domains and projecting onto their intersection using the projection onto convex sets (POCS) algorithm, we obtain an estimate that effectively combines estimates from multiple wavelet domains. While simple, the algorithm provides significant improvement over conventional wavelet shrinkage algorithms based on a single wavelet domain.

**Index Terms**—Besov, denoising, POCS, wavelet.

## I. INTRODUCTION

REMOVING noise from images is an important problem in image acquisition and processing. Among many approaches, denoising by wavelet coefficient thresholding and its variations provide excellent performance for removing noise from photographic images [1]–[3]. Simple and fast, wavelet thresholding suppresses the corrupting noise while effectively preserving image edges, unlike other filtering methods.

Wavelet thresholding is based on the property that typical real-world images have sparse representations in the wavelet domain. Most of the wavelet coefficients have very small magnitudes, while only a few large ones represent the singularities (edges, ridges, etc). Because white noise disperses evenly among all wavelet coefficients, removing small wavelet coefficients suppresses most of the noise energy while preserving most of the image energy.

For maximum denoising performance, we must use the wavelet basis that provides the sparsest representation [1], [3]. In particular, the smoothness of the wavelet transform needs to match the image’s smoothness. However, since different parts of an image have different smoothness, these different parts are most sparsely represented in different wavelet domains. This motivates the use of *multiple* wavelet transform domains to denoise images. However, in conventional wavelet shrinkage,

it is not clear how to take best advantage of multiple wavelet domains.

There have been several previous attempts to use multiple wavelet bases for image denoising. In [4], an estimate from one wavelet domain was used to design a Wiener filter for another wavelet domain. The algorithm in [5] is most similar to our proposed algorithm. The authors define multiple convex sets containing the given noisy data in multiple wavelet domains and use the projections onto convex sets (POCS) algorithm to find the projection onto the intersection. However, although the proposed algorithm provides significantly better denoising results than conventional wavelet thresholding, their convex sets contain the original noisy signal making the performance very sensitive to the initial starting point of the POCS iteration.

In this paper, we develop a multiple wavelet basis denoising algorithm based on the relationship between wavelet thresholding and a variational problem with Besov norm regularization. The algorithm corresponds to POCS with *Besov balls* as the convex sets. The proposed algorithm gives up to 2 dB improvement in PSNR for typical images compared with the best single-basis wavelet shrinkage algorithm.

## II. IMAGE DENOISING BY WAVELET SHRINKAGE

### A. Discrete Wavelet Transform

The discrete wavelet transform (DWT) represents a two-dimensional (2-D) image  $z(x, y)$  in terms of shifted versions of a lowpass scaling function  $\phi(x, y)$  and shifted and dilated versions of three wavelet functions  $\psi^{\text{HL}}(x, y)$ ,  $\psi^{\text{LH}}(x, y)$ , and  $\psi^{\text{HH}}(x, y)$ . Setting  $\Psi := \{\psi^{\text{HL}}, \psi^{\text{LH}}, \psi^{\text{HH}}\}$ , then special choices of the set of functions  $\{\psi_{j,\mathbf{k}} = 2^j \psi(2^j \mathbf{x} - \mathbf{k})\}_{\psi \in \Psi, j \in \mathbf{Z}, \mathbf{k} \in \mathbf{Z}^2}$  and  $\{\phi_{j,\mathbf{k}} = 2^j \phi(2^j \mathbf{x} - \mathbf{k})\}_{j \in \mathbf{Z}, \mathbf{k} \in \mathbf{Z}^2}$  form an orthonormal basis for  $L_2(\mathbf{R}^2)$ . That is, for every  $z \in L_2(\mathbf{R}^2)$ , we have

$$z = \sum_{j_0 \leq j, \mathbf{k} \in \mathbf{Z}^2, \psi \in \Psi} w_{j,\mathbf{k},\psi} \psi_{j,\mathbf{k}} + \sum_{\mathbf{k} \in \mathbf{Z}^2} u_{j_0,\mathbf{k}} \phi_{j_0,\mathbf{k}} \quad (1)$$

with  $w_{j,\mathbf{k},\psi} := \int_{\mathbf{R}^2} z(\mathbf{x}) \psi_{j,\mathbf{k}}(\mathbf{x}) d\mathbf{x}$  and  $u_{j_0,\mathbf{k}} := \int_{\mathbf{R}^2} z(\mathbf{x}) \phi_{j_0,\mathbf{k}}(\mathbf{x}) d\mathbf{x}$ .

With proper prefiltering [6], we can approximate the scaling coefficients of the continuous-space image  $z(x, y)$  at scale  $J$  by the sampled image  $z[\mathbf{k}]$ . That is, the sampled image  $z[\mathbf{k}] = u_{J,\mathbf{k}}$ . Under this approximation, we can compute the coefficients  $w_{j,\mathbf{k},\psi}$  and  $u_{j_0,\mathbf{k}}$  for scales  $j \leq J$ .

Manuscript received August 13, 2003; revised January 19, 2004. This work was supported by NSF Grants MIP-9457438 and CCR-9973188, ONR Grant N00014-02-1-0353, DARPA/AFOSR Grants F49620-97-1-0513 & F49620-01-1-0378, and the Texas Instruments Leadership University Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Paulo Goncalves.

The authors are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: choi@ece.rice.edu; richb@ece.rice.edu).

Digital Object Identifier 10.1109/LSP.2004.833493

## B. Wavelet Thresholding

Image denoising aims to recover an original image from a noisy observation corrupted by additive noise. Let  $z$  be an  $N \times N$  discrete image. We observe the noisy rendition

$$y[\mathbf{k}] = z[\mathbf{k}] + d[\mathbf{k}] \quad (2)$$

where  $d[\mathbf{k}]$  is additive white Gaussian noise of variance  $\sigma^2$ . We aim to estimate  $z[\mathbf{k}]$  from  $y[\mathbf{k}]$ .

Denoising by wavelet shrinkage is based on the property that a typical real-world image  $z$  has a sparse representation in the wavelet domain. Let  $\{\tilde{w}_i\}$  be the DWT coefficients of  $y$ , where  $i = (j, \mathbf{k})$  is an abstract index. To denoise, we shrink each wavelet coefficient by a scalar shrinkage operator  $\theta$  to obtain  $\hat{w}_i = \theta(\tilde{w}_i)$ . The estimate of  $z$  is obtained by computing the inverse wavelet transform of the thresholded wavelet coefficients  $\hat{w}_i$ .

Several different types of thresholding are oft-used in image denoising. Hard-thresholding simply sets the coefficients below a threshold  $T$  to zero. Soft-thresholding first shrinks each coefficient by  $T$  and then hard thresholds [1], [3].

Because wavelet denoising is based on the compaction of signal energy in the wavelet domain, the performance of wavelet shrinkage hinges on the sparsity of the image representation in the wavelet domain. However, since an image usually consists of several regions of different smoothness, the sparsity of its representation in a single wavelet domain is limited. This naturally motivates using multiple wavelet transforms to denoise.

In this paper, we will exploit the link between wavelet shrinkage and a variational problem based on the Besov smoothness norm [7] to develop a multiple wavelet bases denoising algorithm. Before developing the denoising algorithm, we first review some Besov space theory and the link to wavelet shrinkage.

## III. BESOV SPACE AND VARIATIONAL PROBLEMS

### A. Besov Smoothness Spaces

Wavelets provide a simple characterization for a wide variety of function *smoothness spaces* [8]. The norms of these spaces measure image smoothness; smaller norms imply smoother functions. The scale of *Besov spaces*  $B_q^\alpha(L_p)$ ,  $0 < \alpha < \infty$ ,  $0 < p \leq \infty$ ,  $0 < q \leq \infty$ , are particularly useful, for they contain many life-like images. For analyzing wavelet functions possessing  $r > \alpha$  vanishing moments [3], the Besov norm  $\|z\|_{B_q^\alpha(L_p)}$  is equivalent to a sequence norm on the wavelet coefficients of  $z$  [8]

$$\|z\|_{B_q^\alpha(L_p)} \equiv \|u_{j_0, \mathbf{k}}\|_p + \left( \sum_{j \geq j_0} \left( \sum_{\mathbf{k}, \psi} 2^{j(\alpha p + p - 2)} |w_{j, \mathbf{k}, \psi}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}. \quad (3)$$

The three Besov hyperparameters have natural interpretations: a  $p$ -norm of the wavelet coefficients is taken within each scale  $j$ , a weighted  $q$ -norm is taken across scale, and the smoothness parameter  $\alpha$  controls the rate of decay of the  $w_{j, \mathbf{k}, \psi}$  across scale (frequency). Roughly speaking, the parameter  $\alpha$  corresponds to

the number of well-behaved derivatives of  $z$  in  $L_p$ ; hence, the larger the  $\alpha$ , the smoother the functions in  $B_q^\alpha(L_p)$  [8].

Note that any wavelet basis having  $r > \alpha$  vanishing moments can be used to measure a Besov norm [9]. However, depending on the choice of the wavelet, the actual value of the norm will be different.

For discrete images, the wavelet coefficients beyond the finest scale  $J$  are not available, and the Besov norm is defined by limiting the summation in (3) to the available scales.

For image processing applications, we have two particular cases of interest. The first case is when  $p = q$ . Then the Besov norm reduces to

$$\|z\|_{B_p^\alpha(L_p)} \equiv \|u_{j_0, \mathbf{k}}\|_p + \left( \sum_{j_0 \leq j, \mathbf{k}, \psi} 2^{j(\alpha p + p - 2)} |w_{j, \mathbf{k}, \psi}|^p \right)^{\frac{1}{p}} \quad (4)$$

which is simply a weighted  $\ell_p$ -norm of the wavelet coefficients. For example, the  $B_1^1(L_1)$  norm is simply the  $\ell_1$  wavelet norm

$$\|z\|_{B_1^1(L_1)} \equiv \|u_{j_0, \mathbf{k}}\|_1 + \sum_{j_0 \leq j, \mathbf{k}, \psi} |w_{j, \mathbf{k}, \psi}|. \quad (5)$$

A closely related Besov space is  $B_\infty^1(L_1)$ , with norm

$$\|z\|_{B_\infty^1(L_1)} \equiv \|u_{j_0, \mathbf{k}}\|_1 + \sup_{j_0 \leq j, \mathbf{k}, \psi} |w_{j, \mathbf{k}, \psi}|. \quad (6)$$

Note that we have the Besov sandwich  $B_1^1(L_1) \subset BV \subset B_\infty^1(L_1)$  with the set of bounded variation (BV) images an attractive model for real-world images [10].

The Besov norm is closely related to the complexity of an image for image compression purposes [8]. Because the coding complexity under many image compression algorithms is determined by the sparsity of the wavelet coefficients [8], [11], the Besov norm can be interpreted as a measure of wavelet coefficient compactness. An image has small Besov norm if the wavelet coefficients in each scale have small  $\ell_p$  norm and those  $\ell_p$  norms decay rapidly across scale. For small  $p$ , small  $\ell_p$  norm arises when an image's energy is sparsely distributed in the wavelet domain. And since the number of wavelet coefficients increases exponentially across scale, the decay of the  $\ell_p$  norms enforces even further wavelet domain sparsity.

In our denoising algorithm, we will seek an estimate that has small Besov norms in *multiple* wavelet domains to exploit sparse representation in multiple domains.

### B. Link to Wavelet Shrinkage

In [7], it was shown that wavelet shrinkage is closely related to a variational problem with Besov norm regularization. Given an image  $y$ , consider the optimization problem

$$\min_x \|x - y\|_2^2 + \lambda \|x\|_{B_q^\alpha(L_p)}^p. \quad (7)$$

Amazingly, for  $p = q = 1$  and  $\alpha = 1$ , the image  $x$  minimizing (7) is obtained by soft-thresholding with threshold  $T = \lambda/2$ . In this case, the Besov norm is the  $\ell_1$  norm of the coefficients, and the problem in (7) becomes similar to the Basis Pursuit algorithm proposed in [12] where the authors use the  $\ell_1$  norm as a sparsity measure.



Fig. 1. Image denoising using projections onto Besov Balls. (a)  $512 \times 512$  original Lena image with 256 gray levels. (b) Noisy image ( $\sigma = 35$ , PSNR = 17.2 dB). (c) MATLAB's spatially adaptive Wiener filter, `wiener2.m` ( $3 \times 3$  window, PSNR = 27.0 dB). (d) Hard thresholding in  $D_{12}$  domain using best threshold (PSNR = 26.1 dB); the threshold was chosen to maximize PSNR. (e) Besov ball projection in  $B_{\infty}^1(L_1)$  using  $D_{12}$  wavelet (PSNR = 27.2 dB). (f) Besov ball projection in  $B_{\infty}^1(L_1)$  using 5 wavelet bases,  $D_6, D_8, D_{10}, D_{12}$  and  $D_{14}$  (PSNR = 28.3 dB).

For other Besov spaces, such as  $B_2^1(L_1)$  and  $B_{\infty}^1(L_1)$ , the solution of the variational problem (7) corresponds to scale-dependent soft-thresholding [7]. Noting that the Besov norm measures the wavelet transform's compactness, the optimization (7) can be interpreted as image estimation under a wavelet sparsity penalty.

The minimization of (7) is equivalent to the dual constrained optimization problem

$$\min_x \|x - y\|_2 \text{ subject to } \|x\|_{B_q^{\alpha}(L_p)} \leq r \quad (8)$$

where  $r$  is a parameter determined by  $\lambda$ . The solution to this constrained optimization is the  $\ell_2$  projection of  $y$  onto the Besov ball of radius  $r$

$$B(r) \equiv \left\{ x : \|x\|_{B_q^{\alpha}(L_p)} \leq r \right\}. \quad (9)$$

For  $p, q \geq 1$ , the Besov ball is convex, and thus a unique projection always exists. Projecting a noisy image onto a Besov ball finds an estimate that is close to the original noisy image but with a sparse representation in the wavelet domain defining the Besov ball.

When  $p = q$ , Besov balls specify constraints on the  $\ell_p$  norm of wavelet coefficients. In [13], similar constraints on the moments of the wavelet coefficients were defined to derive the maximum entropy prior to be used in multiple wavelet domain maximum a posteriori (MAP) estimation.

#### IV. MULTIPLE WAVELET BASIS DENOISING

Since wavelet shrinkage finds an image with a sparse wavelet representation, it is natural to wish and try to be sparse in multiple wavelet domains simultaneously in order to properly

handle image regions with differing smoothness. We rely on the key fact that the structure of a Besov ball varies slightly as we change the wavelet basis used to compute the Besov norm. Thus, in our multiple basis image denoising algorithm, we will project alternately onto the Besov balls defined by the same Besov parameters but *different wavelet bases*.

Given  $M$  different wavelet bases, define the corresponding Besov balls  $B_i(r), i = 1, \dots, M$ . We wish to find the multiple wavelet basis denoising estimate  $x \in \cap_{i=1}^M B_i(r_i)$  that is close to the given noisy image  $y$ .

Since each Besov ball  $B_i(r_i)$  is a convex set, we can find an estimate using the *projections onto convex sets* (POCS) algorithm. We simply project onto each Besov ball in turn until the algorithm converges. This algorithm is guaranteed to converge to a point in the intersection, which we know exists.

For general  $B_q^{\alpha}(L_p)$ , computing the projection requires solving a constrained optimization problem. However, for the Besov balls in  $B_{\infty}^1(L_1)$ , we can compute the projections by simply projecting onto hyperplanes. Assuming that we do not change the scaling coefficients  $u_{j_0, \mathbf{k}}$ , the projection onto  $B_i(r_i)$  corresponds to projecting the wavelet coefficients  $\bar{w}_{j, \mathbf{k}, \psi}$  in each scale onto the set  $\sum_{\mathbf{k}, \psi} |w_{j, \mathbf{k}, \psi}| \leq r_i - \|u_{j_0, \mathbf{k}}\|_1$ . This corresponds to solving the following general projection problem in  $\mathbb{R}^n$  at each scale

$$\min_{w \in \mathbb{R}^n} \|w - \bar{w}\|_2 \text{ subject to } \|w\|_1 \leq C. \quad (10)$$

Clearly, if  $\bar{w}$  is in the set  $\|w\|_1 \leq C$ , then the solution is  $w = \bar{w}$ . If not, then the solution is easily found by first recognizing that either  $\text{sgn}(\bar{w}_i) = \text{sgn}(w_i)$  or  $w_i = 0$ ; otherwise, we could find another  $w$  closer to  $\bar{w}$  by flipping the signs of the  $w_i$ 's so that  $\text{sgn}(w_i) = \text{sgn}(\bar{w}_i)$ . Now, knowing the signs of the

$w_i$ 's, we can find the orthogonal projection  $\tilde{w}$  of  $\bar{w}$  onto the hyperplane  $\text{sgn}(\bar{w}_i)w_i = C$ . The solution to (10) is then the point on the same hyperplane closest to  $\tilde{w}$  with the correct signs. If  $\text{sgn}(\tilde{w}_i) \neq \text{sgn}(\bar{w}_i)$ , then we realize that  $w_i$  must be zero, and we reformulate the problem (10) by fixing those  $w_i$ 's that are zero and repeating the process until all  $w_i$ 's are determined. This process is guaranteed to terminate in a finite number of steps.

Because the projection computations are straightforward in each wavelet domain, the majority of the computational cost of the algorithm results from the computation of the multiple wavelet transforms. In particular, because we need to alternately project onto Besov balls in different wavelet domains, we must perform  $M$  forward and inverse wavelet transforms per POCS iteration. Thus, for  $K$  POCS iterations, the total computational cost is  $O(KMN \log N)$  for an  $N$ -pixel image.

For proper operation, we must set the radii of the Besov balls properly. Setting the radius of each ball is equivalent to choosing the threshold in wavelet thresholding. Although there has been some effort to set the threshold optimally [2], properly setting it to obtain good denoising results is a very delicate problem. When the noise level is moderate, our experiments suggest that we should set the radii close to the actual Besov norm of the (unknown) original image but somewhat larger to include it in the ball. However, since the original image is unknown, this is impossible. A more practical approach first obtains a rough pilot estimate and computes its Besov norm. Then, to guarantee with high probability that the original image is within the balls, we set their radii somewhat larger than this estimated norm.

## V. NUMERICAL RESULTS

To demonstrate our proposed scheme, we added white Gaussian noise to a number of test images and then estimated the original image from the noisy version. Thanks to its close relationship with BV, we focus on  $B_\infty^1(L_1)$  Besov balls. To set the radii, we obtained a pilot estimate using hard-thresholding in one wavelet domain and then computed its Besov norm in a different wavelet domain. We then set the radii of the Besov balls to 1.2 times this estimated norm.

Fig. 1 provides an example for the  $512 \times 512$  Lena image corrupted with white Gaussian noise of  $\sigma = 35$ . We compare our algorithm with simple wavelet hard thresholding and MATLAB's spatially adaptive Wiener filter (wiener2.m). We also show the result of a Besov ball projection using only one wavelet basis. For the multiple basis algorithm, we used five Daubechies' orthogonal wavelets  $D_6, D_8, D_{10}$ , and  $D_{12}$  with differing number of vanishing moments. Multiple Besov ball denoising outperforms the single basis method by over 1 dB in peak-signal-to-noise-ratio (PSNR  $\equiv 10 \log(255^2 / \|z - \hat{z}\|_2^2)$ ).

Table I details the denoising performance of our algorithm for several different example images and two different noise variances. For the multiple Besov ball method, we used the same five wavelet bases as in Fig. 1. For all images and all noise levels, our algorithm outperforms conventional wavelet hard thresholding in the  $D_{12}$  domain by as much as 2–2.5 dB in PSNR.

TABLE I

PSNR's OF DENOISING RESULTS FOR SEVERAL IMAGES AT  $\sigma = 25$  AND 35. WE COMPARE THE PERFORMANCE OF WAVELET HARD THRESHOLDING (WHT), MATLAB'S SPATIALLY VARYING WIENER FILTER (wiener2.m), AND THE PROPOSED BESOV BALL PROJECTION (BB).

Image	$\sigma_n$	Noisy	WHT	wiener2	BB
Lena (512 × 512)	25	20.16	27.12	27.23	29.41
	35	17.27	25.73	24.40	28.01
Barbara (512 × 512)	25	20.17	24.08	25.47	26.72
	35	17.24	22.85	23.21	24.38
Boat (512 × 512)	25	20.18	26.24	26.95	28.55
	35	17.25	24.92	24.27	27.00
Bridge (256 × 256)	25	20.15	22.35	24.83	24.38
	35	17.27	21.44	18.19	23.06
Cameraman (256 × 256)	25	20.16	24.02	26.24	26.32
	35	17.25	22.68	23.71	24.79
Fruit (256 × 256)	25	20.18	26.11	27.51	28.50
	35	17.27	24.59	24.56	27.03

## VI. CONCLUSIONS AND FUTURE RESEARCH

We have proposed an image denoising algorithm based on projecting a noisy image onto Besov balls in multiple wavelet transform domains. While simple, this approach can improve significantly over standard methods such as wavelet hard thresholding in a single wavelet domain. While we developed and demonstrated the algorithm with 2-D images, its application to 1-D or multidimensional signal denoising is straightforward. Interesting topics for future investigation include how to optimally set the radii of the Besov balls and how to quantify their effect on the resulting estimate.

## REFERENCES

- [1] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [2] —, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, pp. 1200–1224, Dec. 1995.
- [3] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.
- [4] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical Wiener filtering," *Proc. SPIE*, vol. 3169, pp. 389–399, July 1997.
- [5] P. Ishwar, K. Ratakonda, P. Moulin, and N. Ahuja, "Image denoising using multiple compaction domains," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc.—ICASSP'98*, Seattle, WA, May 1998, pp. 1889–1892.
- [6] P. Abry and P. Flandrin, "On the initialization of the discrete wavelet transform algorithm," *IEEE Signal Processing Lett.*, vol. 1, no. 2, pp. 32–34, Feb. 1994.
- [7] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Processing*, vol. 7, pp. 319–355, Mar. 1998.
- [8] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, Mar. 1992.
- [9] Y. Meyer, *Wavelets and Operators*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [10] S. Osher, L. I. Rudin, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.
- [11] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, "Tree approximation and optimal encoding," *J. Appl. Comput. Harmon. Anal.*, vol. 11, no. 2, pp. 192–226, Sept. 2001.
- [12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," in *SIAM J. Sci. Comput.*, vol. 20, 1999, pp. 33–61.
- [13] P. Ishwar and P. Moulin, "Multiple-domain image modeling and restoration," in *Proc. IEEE Int. Conf. Image Processing '99*, Kobe, Japan, Oct. 1999, pp. 362–336.